

0590
0731

13



OIIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/964,275B

DATE: 08/07/2002

TIME: 09:50:02

Input Set : A:\0136542 SEQLST.ST25.txt

Output Set: N:\CRF3\08072002\I964275B.raw

```

3 <110> APPLICANT: Dai, Ken-Shwo
4      Yang, Chin-Chang
6 <120> TITLE OF INVENTION: HUMAN NUC2-RELATED GENE VARIANTS ASSOCIATED WITH LUNG CANCER
8 <130> FILE REFERENCE: U 013654-2
10 <140> CURRENT APPLICATION NUMBER: 09/964,275B
11 <141> CURRENT FILING DATE: 2001-09-26
13 <160> NUMBER OF SEQ ID NOS: 10
15 <170> SOFTWARE: PatentIn version 3.1
17 <210> SEQ ID NO: 1
18 <211> LENGTH: 2385
19 <212> TYPE: DNA
20 <213> ORGANISM: HOMO SAPIEN
22 <220> FEATURE:
23 <221> NAME/KEY: CDS
24 <222> LOCATION: (145)..(1032)
25 <223> OTHER INFORMATION:
27 <400> SEQUENCE: 1
28 ggctctcat ctggaacacc tcgggtcacc cccgacaacg gtggtgggag ggagagcggc      60
30 ctctctctcc ctggtggggc ctgtctgggt gaagcccctc tgttcccag gatcgtccca      120
32 acccccagcc ggggtgctcg agcc atg gcc gac acc atc ttc ggc agc ggg      171
33                               Met Ala Asp Thr Ile Phe Gly Ser Gly
34                               1           5
36 aat gat cag tgg gtt tgc ccc aat gac cgg cag ctt gcc ctt cga gcc      219
37 Asn Asp Gln Trp Val Cys Pro Asn Asp Arg Gln Leu Ala Leu Arg Ala
38 10                               15           20           25
40 aag ctg cag acg ggc tgg tcc gtg cac acc tac cag acg gag aag cag      267
41 Lys Leu Gln Thr Gly Trp Ser Val His Thr Tyr Gln Thr Glu Lys Gln
42                               30           35           40
44 agg agg aag cag cac ctc agc ccg gcg gag gtg gag gcc atc ctg cag      315
45 Arg Arg Lys Gln His Leu Ser Pro Ala Glu Val Glu Ala Ile Leu Gln
46                               45           50           55
48 gtc atc cag agg gca gag cgg ctc gac gtc ctg gag cag cag aga atc      363
49 Val Ile Gln Arg Ala Glu Arg Leu Asp Val Leu Glu Gln Gln Arg Ile
50                               60           65           70
52 ggg cgg ctg gtg gag cgg ctg gag acc atg agg cgg aat gtg atg ggg      411
53 Gly Arg Leu Val Glu Arg Leu Glu Thr Met Arg Arg Asn Val Met Gly
54                               75           80           85
56 aac ggc ctg tcc cag tgt ctg ctc tgc ggg gag gtg ctg ggc ttc ctg      459
57 Asn Gly Leu Ser Gln Cys Leu Leu Cys Gly Glu Val Leu Gly Phe Leu
58 90                               95           100           105
60 ggc agc tcg tcg gtg ttc tgc aaa gac tgc agg aag gtc tgg aag agg      507
61 Gly Ser Ser Ser Val Phe Cys Lys Asp Cys Arg Lys Val Trp Lys Arg
62                               110           115           120

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/964,275B

DATE: 08/07/2002

TIME: 09:50:02

Input Set : A:\0136542 SEQLST.ST25.txt

Output Set: N:\CRF3\08072002\I964275B.raw

64	tcg ggg gcc tgg ttc tac aaa ggg ctc ccc aag tat atc ttg ccc ctg	555
65	Ser Gly Ala Trp Phe Tyr Lys Gly Leu Pro Lys Tyr Ile Leu Pro Leu	
66	125 130 135	
68	aag acc cct ggc cga gct gat gag ccc cag ttc cga cct tgg ccc acg	603
69	Lys Thr Pro Gly Arg Ala Asp Glu Pro Gln Phe Arg Pro Trp Pro Thr	
70	140 145 150	
72	gaa ccg gca gag cga gag ccc aga agc tct gag acc agc cgc atc tac	651
73	Glu Pro Ala Glu Arg Glu Pro Arg Ser Ser Glu Thr Ser Arg Ile Tyr	
74	155 160 165	
76	acg tgg gcc cga gga aga gtg gtt tcc agt gac agt gac agt gac tcg	699
77	Thr Trp Ala Arg Gly Arg Val Val Ser Ser Asp Ser Asp Ser Asp Ser	
78	170 175 180 185	
80	gat ctt agc tcc tcc agc cta gag gac aga ctc cca tcc act ggg gtc	747
81	Asp Leu Ser Ser Ser Leu Glu Asp Arg Leu Pro Ser Thr Gly Val	
82	190 195 200	
84	agg gac ccg aaa ggc gac aaa ccc tgg aag gag tca ggt ggc agc gtg	795
85	Arg Asp Arg Lys Gly Asp Lys Pro Trp Lys Glu Ser Gly Gly Ser Val	
86	205 210 215	
88	gag gcc ccc agg atg ggg ttc acc caa ccc gcg ggc cac ctc ttt ggg	843
89	Glu Ala Pro Arg Met Gly Phe Thr Gln Pro Ala Gly His Leu Phe Gly	
90	220 225 230	
92	ttg cag agc agc ctg gcc agt ggt gag acg ggc aca ggc tct gct gac	891
93	Leu Gln Ser Ser Leu Ala Ser Gly Glu Thr Gly Thr Gly Ser Ala Asp	
94	235 240 245	
96	ccg cca ggg gga ggg aca ggc tct gct gac ccg cca ggg gga ccc cgc	939
97	Pro Pro Gly Gly Gly Thr Gly Ser Ala Asp Pro Pro Gly Gly Pro Arg	
98	250 255 260 265	
100	ccc ggg ctg acc cga agg gcc ccg gta aaa gac aca cct gga cga gcc	987
101	Pro Gly Leu Thr Arg Arg Ala Pro Val Lys Asp Thr Pro Gly Arg Ala	
102	270 275 280	
104	ccc gct gct gac gca gct cca gca ggc ccc tcc agc tgc ctg ggc	1032
105	Pro Ala Ala Asp Ala Ala Pro Ala Gly Pro Ser Ser Cys Leu Gly	
106	285 290 295	
108	tgaggtgtct ggtgcctgga acagaattcc ctgtggagga ttcttgccag accctgccc	1092
110	gctcctccct gaccggtcct tgtgccctca ccagacacc tggtggccat gactcaacc	1152
112	accagtgttg ggagccgtct gcctccccag ctcaagtgcct ttctgcacc cttctctcct	1212
114	ggggagctgt ctgcatccgc caccctctcc aaccactgcc ctccagcccc gacctattt	1272
116	attaccctcc cctccacac ccccaatcta cctggtgatg attttaagtt tgcgcgtgtc	1332
118	ttgggttggg ctggggggtt tcccacatgc agtgtcagag gggccgccc gtggggctat	1392
120	ctccgttgct atattaatgg caagactaaa tgaaacctag ggcaaggcct ccgaagctgc	1452
122	gtgtggcccc ttagaggtga gcatcagagc cagagcagtg agggggagac tcaccacccc	1512
124	tctccctctc ccttcagctc tgggaggcag gcgcagtgcc cccctcccat gggctggccc	1572
126	aggaccgcgg gtgaaacctg ggtctgttta gtttctttg tttttgtatg tttgtttgtt	1632
128	tttgacacag tctcgctttg ttgcccaggc tggggtgcag tggcacgatc gcggctcact	1692
130	gcaacctcca cctcccgggc tcaagcgatt ctctcacctc agcctcctga gtagggtgga	1752
132	ttacagatgc ccgccaccac acccagttaa tttttgtatt tttagaagag atgggggttc	1812
134	tccatgttg ccaggtggt cttgaactcc tgggtctcaag tgatccgccc gcctcggcct	1872
136	cccaaagtgc tgggattaca ggtgtgagcc accgcaccca atcctattag gtttctttga	1932
138	atccctcat ggcctgctg gtttttgcctc agcctgtctt cagcttgagg agctgggaag	1992

RAW SEQUENCE LISTING

DATE: 08/07/2002

PATENT APPLICATION: US/09/964,275B

TIME: 09:50:02

Input Set : A:\0136542 SEQLST.ST25.txt

Output Set: N:\CRF3\08072002\I964275B.raw

```

140 ctctggtgga tgctatgaac tcacttgctg aagagcagcg ttcaggtgca tccccagcca 2052
142 gggcacgtgg ctcctcagc catgaattca cttctcttca ggagggttg cttggcatga 2112
144 aaatacttca ttcagagtat gggcaaagtc ttctggaaaa cccttccttg aagagagaga 2172
146 acgtgtgtgt gtgtgtcggg gatcacaccc tcccatcctt cctgcctcct gccccaaacc 2232
148 cggggttcct ggggtctggaa gggccttctc tccaagctgg gagctcctgg gccccacca 2292
150 ttcacttttt gtccttgctg ctggcaaaaca gtaaagaaac tcactttccc tgtggcacgt 2352
152 tatgcttcag aattaaaaca atgaagatta aaa 2385
155 <210> SEQ ID NO: 2
156 <211> LENGTH: 296
157 <212> TYPE: PRT
158 <213> ORGANISM: HOMO SAPIEN
160 <400> SEQUENCE: 2
162 Met Ala Asp Thr Ile Phe Gly Ser Gly Asn Asp Gln Trp Val Cys Pro
163 1 5 10 15
166 Asn Asp Arg Gln Leu Ala Leu Arg Ala Lys Leu Gln Thr Gly Trp Ser
167 20 25 30
170 Val His Thr Tyr Gln Thr Glu Lys Gln Arg Arg Lys Gln His Leu Ser
171 35 40 45
174 Pro Ala Glu Val Glu Ala Ile Leu Gln Val Ile Gln Arg Ala Glu Arg
175 50 55 60
178 Leu Asp Val Leu Glu Gln Gln Arg Ile Gly Arg Leu Val Glu Arg Leu
179 65 70 75 80
182 Glu Thr Met Arg Arg Asn Val Met Gly Asn Gly Leu Ser Gln Cys Leu
183 85 90 95
186 Leu Cys Gly Glu Val Leu Gly Phe Leu Gly Ser Ser Ser Val Phe Cys
187 100 105 110
190 Lys Asp Cys Arg Lys Val Trp Lys Arg Ser Gly Ala Trp Phe Tyr Lys
191 115 120 125
194 Gly Leu Pro Lys Tyr Ile Leu Pro Leu Lys Thr Pro Gly Arg Ala Asp
195 130 135 140
198 Glu Pro Gln Phe Arg Pro Trp Pro Thr Glu Pro Ala Glu Arg Glu Pro
199 145 150 155 160
202 Arg Ser Ser Glu Thr Ser Arg Ile Tyr Thr Trp Ala Arg Gly Arg Val
203 165 170 175
206 Val Ser Ser Asp Ser Asp Ser Asp Ser Asp Leu Ser Ser Ser Ser Leu
207 180 185 190
210 Glu Asp Arg Leu Pro Ser Thr Gly Val Arg Asp Arg Lys Gly Asp Lys
211 195 200 205
214 Pro Trp Lys Glu Ser Gly Gly Ser Val Glu Ala Pro Arg Met Gly Phe
215 210 215 220
218 Thr Gln Pro Ala Gly His Leu Phe Gly Leu Gln Ser Ser Leu Ala Ser
219 225 230 235 240
222 Gly Glu Thr Gly Thr Gly Ser Ala Asp Pro Pro Gly Gly Gly Thr Gly
223 245 250 255
226 Ser Ala Asp Pro Pro Gly Gly Pro Arg Pro Gly Leu Thr Arg Arg Ala
227 260 265 270
230 Pro Val Lys Asp Thr Pro Gly Arg Ala Pro Ala Ala Asp Ala Ala Pro
231 275 280 285
234 Ala Gly Pro Ser Ser Cys Leu Gly

```

RAW SEQUENCE LISTING

DATE: 08/07/2002

PATENT APPLICATION: US/09/964,275B

TIME: 09:50:02

Input Set : A:\0136542 SEQLST.ST25.txt

Output Set: N:\CRF3\08072002\I964275B.raw

```

235      290      295
238 <210> SEQ ID NO: 3
239 <211> LENGTH: 2472
240 <212> TYPE: DNA
241 <213> ORGANISM: HOMO SAPIEN
243 <220> FEATURE:
244 <221> NAME/KEY: CDS
245 <222> LOCATION: (145)..(1119)
246 <223> OTHER INFORMATION:
248 <400> SEQUENCE: 3
249 ggctcctcat ctggaacacc tcgggtcacc cccgacaacg gtgggtgggag ggagagcggc      60
251 ctctctctcc ctggtggggc ctgtctgggt gaagcccctc tgttcccag gatcgtccca      120
253 acccccagcc ggggtgctccg agcc atg gcc gac acc atc ttc ggc agc ggg      171
254                               Met Ala Asp Thr Ile Phe Gly Ser Gly
255                               1                               5
257 aat gat cag tgg gtt tgc ccc aat gac cgg cag ctt gcc ctt cga gcc      219
258 Asn Asp Gln Trp Val Cys Pro Asn Asp Arg Gln Leu Ala Leu Arg Ala
259 10                               15                               20                               25
261 aag ctg cag acg ggc tgg tcc gtg cac acc tac cag acg gag aag cag      267
262 Lys Leu Gln Thr Gly Trp Ser Val His Thr Tyr Gln Thr Glu Lys Gln
263                               30                               35                               40
265 agg agg aag cag cac ctc agc ccg gcg gag gtg gag gcc atc ctg cag      315
266 Arg Arg Lys Gln His Leu Ser Pro Ala Glu Val Glu Ala Ile Leu Gln
267                               45                               50                               55
269 gtc atc cag agg gca gag cgg ctc gac gtc ctg gag cag cag aga atc      363
270 Val Ile Gln Arg Ala Glu Arg Leu Asp Val Leu Glu Gln Gln Arg Ile
271                               60                               65                               70
273 ggg cgg ctg gtg gag cgg ctg gag acc atg agg cgg aat gtg atg ggg      411
274 Gly Arg Leu Val Glu Arg Leu Glu Thr Met Arg Arg Asn Val Met Gly
275                               75                               80                               85
277 aac ggc ctg tcc cag tgt ctg ctc tgc ggg gag gtg ctg ggc ttc ctg      459
278 Asn Gly Leu Ser Gln Cys Leu Leu Cys Gly Glu Val Leu Gly Phe Leu
279 90                               95                               100                               105
281 ggc agc tcg tcg gtg ttc tgc aaa gac tgc agg aag aaa gtc tgc acc      507
282 Gly Ser Ser Ser Val Phe Cys Lys Asp Cys Arg Lys Lys Val Cys Thr
283                               110                               115                               120
285 aaa tgt ggg atc gag gcc tcc cct ggc cag aag cgg ccc ctg tgg ctg      555
286 Lys Cys Gly Ile Glu Ala Ser Pro Gly Gln Lys Arg Pro Leu Trp Leu
287                               125                               130                               135
289 tgt aag atc tgc agt gag caa aga gag gtc tgg aag agg tcg ggg gcc      603
290 Cys Lys Ile Cys Ser Glu Gln Arg Glu Val Trp Lys Arg Ser Gly Ala
291                               140                               145                               150
293 tgg ttc tac aaa ggg ctc ccc aag tat atc ttg ccc ctg aag acc cct      651
294 Trp Phe Tyr Lys Gly Leu Pro Lys Tyr Ile Leu Pro Leu Lys Thr Pro
295                               155                               160                               165
297 ggc cga gct gat gac ccc cac ttc cga cct ttg ccc acg gaa ccg gca      699
298 Gly Arg Ala Asp Asp Pro His Phe Arg Pro Leu Pro Thr Glu Pro Ala
299 170                               175                               180                               185
301 gag cga gag ccc aga agc tct gag acc agc cgc atc tac acg tgg gcc      747

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/964,275B

DATE: 08/07/2002

TIME: 09:50:02

Input Set : A:\0136542 SEQLST.ST25.txt

Output Set: N:\CRF3\08072002\I964275B.raw

```

302 Glu Arg Glu Pro Arg Ser Ser Glu Thr Ser Arg Ile Tyr Thr Trp Ala
303          190          195          200
305 cga gga aga gtg gtt tcc agt gac agt gac agt gac tcg gat ctt agc      795
306 Arg Gly Arg Val Val Ser Ser Asp Ser Asp Ser Asp Ser Asp Leu Ser
307          205          210          215
309 tcc tcc agc cta gag gac aga ctc cca tcc act ggg gtc agg gac cgg      843
310 Ser Ser Ser Leu Glu Asp Arg Leu Pro Ser Thr Gly Val Arg Asp Arg
311          220          225          230
313 aaa ggc gac aaa ccc tgg aag gag tca ggt ggc agc gtg gag gcc ccc      891
314 Lys Gly Asp Lys Pro Trp Lys Glu Ser Gly Gly Ser Val Glu Ala Pro
315          235          240          245
317 agg atg ggg ttc acc caa ccc gcg ggc cac ctc ttt ggg ttg cag agc      939
318 Arg Met Gly Phe Thr Gln Pro Ala Gly His Leu Phe Gly Leu Gln Ser
319 250          255          260          265
321 agc ctg gcc agt ggt gag acg ggc aca ggc tct gct gac ccg cca ggg      987
322 Ser Leu Ala Ser Gly Glu Thr Gly Thr Gly Ser Ala Asp Pro Pro Gly
323          270          275          280
325 gga ggg aca ggc tct gct gac ccg cca ggg gga ccc cgc ccc ggg ctg      1035
326 Gly Gly Thr Gly Ser Ala Asp Pro Pro Gly Gly Pro Arg Pro Gly Leu
327          285          290          295
329 acc cga agg gcc ccg gta aaa gac aca cct gga cga gcc ccc gct gct      1083
330 Thr Arg Arg Ala Pro Val Lys Asp Thr Pro Gly Arg Ala Pro Ala Ala
331          300          305          310
333 gac gca gct cca gca ggc ccc tcc agc tgc ctg ggc tgaggtgtct      1129
334 Asp Ala Ala Pro Ala Gly Pro Ser Ser Cys Leu Gly
335          315          320          325
337 ggtgcctgga acagaattcc ctgtggagga ttctgtccag accctgcccg gctcctccct      1189
339 gaccggtcct tgtgccctca ccagacaccc tggtggccat gactcaacaa accagtgttg      1249
341 ggagccgtct gcctcccag ctcagtgcct ttctgcaccc cttctctcct ggggagctgt      1309
343 ctgcatccgc cccccctcc aaccactgcc ctcagccccc gacctatttt attaccctcc      1369
345 cctcccacac ccccaatcta cctggtgatg attttaagtt tgcgcgtgtc ttgggttggg      1429
347 ctgggggggt tcccacatgc agtgtcagag gggccgcccg gtggggctat ctccgttgct      1489
349 atattaatgg caagactaaa tgaaacctag ggcacggcct ccgaagctgc gtgtggcccc      1549
351 ttagaggtga gcatcagagc cagagcagtg agggggagac tcacccaccc tctccctctc      1609
353 ccttcagctc tgggaggcag gcgcagtgcc cccctcccat gggctggccc aggaccgagg      1669
355 gtgaaacctg ggtctgttta gttcttttg tttttgtatg tttgtttgtt tttgacacag      1729
357 tctcgttttg ttgcccaggc tggggtgcag tggcacgata gcggctcact gcaacctcca      1789
359 cctcccgggc tcaagcgatt ctctcacctc agcctcctga gtagggtgga ttacagatgc      1849
361 ccgccaccac acccagttaa tttttgtatt tttagaagag atgggggttt tccatgttgg      1909
363 ccaggctggt cttgaactcc tggctcgaag tgatccgccc gcctcgccct cccaaagtgc      1969
365 tgggattaca ggtgtgagcc accgcaccca atoctattag gtttctttga atccccctcat      2029
367 ggctgtcctg gtttttgctc agcctgtctt cagcttgagg agctgggaag ctctggtgga      2089
369 tgctatgaac tcacttgctg aagagcagcg ttcagggtgca tcccagcca gggcacgtgg      2149
371 ctccctcagc catgaattca cttctcttca ggaggttttg cttggcatga aaatacttca      2209
373 ttcagagtat gggcaaatgc ttctggaaaa cccttccctg aagagagaga acgtgtgtgt      2269
375 gtgtgtcggg gatcacaccc tcccatcctt cctgcctcct gcccacaaacc ccgggttccct      2329
377 ggggtctgga gggccttctc tccaagctgg gagctcctgg gcccacacca ttcacttttt      2389
379 gtccttgctg ctggcacaaca gtaaagaaac tcactttccc tgtggcacgt tatgtttcag      2449
381 aattaaaaca atgaagatta aaa      2472

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/964,275B

DATE: 08/07/2002

TIME: 09:50:03

Input Set : A:\0136542 SEQLST.ST25.txt

Output Set: N:\CRF3\08072002\I964275B.raw